Overall, the protocol is well-constructed and thorough. A few specific comments:

Enrollment, exposure determination, and cohort assignment

1. An 80% participation rate is optimistic. The investigators might consider how a substantially lower participation rate will affect implementing the study.
2. The parameters for assignment of low-exposure to the active or passive follow-up cohorts should be explicit. Indicating that participants "will be scrutinized to determine whether to enroll them into the active or passive follow-up cohort" is insufficient.
3. The exposure determination is the weakest part of the protocol. Outcomes attributed to exposure to oil (or related materials) will be suspect if the definition of "exposed" as it stands now is used. A "definitely exposed" cohort—perhaps consisting of persons who had visible oil on skin—could make up a biologically-plausible exposure cohort. Even among petroleum worker studies, finding effect from oil exposure has proven tricky. For example, there are relatively more esophageal cancers among roughnecks—but a higher rate of tobacco and alcohol use too.
4. Among the special populations to be recruited/considered are "persons with reactive airway disease." "Reactive airway disease" lacks recognized criteria—it is more of a euphemism than a diagnostic category. If the investigators would like to enroll persons with asthma, chronic obstructive pulmonary disease, and allergic rhinitis, these disease categories should be specified.

Statistics/Epidemiologic analysis:

I understand that recently, regression models and other advanced statistical techniques are used from the get-go (rather than starting with old-fashioned bivariate analysis—how about a 2X2 table?). If one starts out with a regression model, then at least describing how the variables to include in the model are determined should be described: perhaps by outcomes from existing studies? Significance in the bivariate analysis (or in lieu of that, at least the criteria for, say, stepwise determination)?.

Field implementation:

1. Some consideration to the safety of the home-visit interviewers should be discussed. In some settings, training the interviewer teams in situational awareness is sufficient.
2. How interviewers should handle markedly abnormal symptoms or signs (blood pressure 220/120; mention of suicide, etc.) needs mention. Once you're above a few hundred participants, this will happen at least a couple times.
3. "Fainting during blood collection is exceedingly rare." This wording is too strong—it is uncommon, but still happens in 0.25% to 1% of venipunctures (and highest in males <40 years of age—likely common in this cohort). At least word it less strongly, and let the interviewers know how to handle it.
4. Procedures for lost or inadequate samples: At least mention that biologic specimens will either be recollected, or how the partial data will be handled.
5. Data protection: the interviewers' laptops/devices should be encrypted in real time. Anything less is below "standard of care" in 2010.